An Investigation of Testing Methodologies of Second Language Perception

Japanese Listeners' Perception of /r/-/l/ Contrast *

Megumi TAKENAKA **

キーワード： L2 perception, testing methodologies, categorial discrimination test

第二音声知覚研究において、学習者の第二音声知覚能力を評価、数値化する方法がその後の研究の方向性を決めるといってよい。伝統的な同定テスト（identification test）は長期記憶された音韻素性に、その後開発された弁別テストは短期記憶された音響的情報に基づいて音声を判断すると知られており、互いの欠点を補い合うよう両テストの併用が主流となっていた。だが、Flege et al. (1999) は両テストの要素を持つ categorial discrimination test (CDT) を考案することで従来の知覚テスト手法に新手法を提案した。第二音声知覚研究者は次第にこの手法を用いて実験を行うようになってきているが、研究例が多いとは言いがたい。また、Flege et al. (1999) も含め CDT の正当性について 2 テストと比較した研究はまだなく、後続の研究は彼らの主張をそのまま受け入れている。本研究は CDT の有効性を同定・弁別テストと比較し、Flege et al.の主張が正当であるかを改めて確認した。そのうえで音声対立の中で一番知覚が難しいとされる語頭子音群内の/r/-/l/対立音声知覚を調べ、先行研究の結果と比べた。CDT スコアは同定・弁別テストのスコアとの間に著しい差が認められない[$F(12, 26) = 0.14$; $p > 0.40$]ことから、両テストとほぼ同等の結果が出せたと証明された。また同定テストと強い相関関係[$r=0.76$]を示し、音声知覚が音韻素性に基づいて行われているとみなすことができる。だが 2 要因の ANOVA（テストタイプ×ペア）では差が見られ[$F(2, 6) = 14.28$, $p<.006$]、その原因は弁別テストでの/pr/-/pl/対立の音響的特徴（周波数に差がなく、持続時間に差がある）であり、語頭子音群の音声対立を弁別テストで調査するのは不適切と思われる。ただし CDT・同定テストではその影響が見られないので持続時間が L2 知覚に与える影響は弱い。また全ペアの周波数で変化に乏しいことが、F2 を判断基準としている日本人学習者の誤聴の原因となっていると本研究では推測した。加えて、先行研究で/r/>/l/となっている子音群の音声対立知覚が、/fr/-/fl/では成立するものの/pr/-/pl/では反対であった。語頭子音群全般にこのような現象が起こるのか今後さらなる研究が

---

* 第二音声知覚のテスト手法—日本人学習者の/r/-/l/対立音声知覚を対象とした研究—（竹中 芽久美）

** 大阪大学大学院言語文化研究科博士後期課程

必要である。

1. Introduction

One of the most fundamental and important issues in investigating the process of second language (L2) perceptual learning is the stability of perception test. Without an appropriate phonetic evaluation system from a longitudinal perspective, we cannot know whether the perceptual training is suitable for L2 teaching and learning or not. Two kinds of perception tests are often conducted in order to evaluate L2 perceptual learning. Most phonetic researchers use both identification and discrimination tests which are thought to be necessary to yield stable results of perception experiments. The meaning of 'identification' is to label particular sounds with a stable linguistic category, and 'discrimination' means to distinguish the sound that can be perceived as the same as the previous sound when it is demonstrated again in speech perception research. Identification is an absolute judgment, whereas on the contrary, discrimination is a relative judgment. These two types of perception tests are based on different dimensions of speech perception, and therefore researchers needed both of their results to analyze the L2 perception system. However, Flege et al. (1999) designed a categorial discrimination test (CDT) that uses tokens produced by at least three speakers to investigate the phonetic abilities of listeners which can be assessed in both identification and discrimination tests. This study focuses on the 'famous' /r/-/l/ phonetic contrast, especially in the clusters in word-initial position contained in minimal pairs, and examines whether the results of three kinds of perception tests on native speakers of Japanese do not disagree with each other.

2. Cross-linguistic studies of liquids

The difficulty that native Japanese speakers have in perceiving or producing the English /r/-/l/ distinction is a well-known example in phonetic research. The Japanese /r/ is phonetically an alveolar flap that is quite different from American English /r/-/l/ contrast, which is a rhotic (/r/) versus lateral (/l/) contrast. Japanese listeners who have had little conversational experience with native English speakers tend to perceptually assimilate both English liquids to the Japanese /r/ despite the articulatory difference between them (Goto, 1971; Miyawaki et al., 1975; MacKain et al., 1981; Sato, 1984; Broselow & Finer, 1991; Eckman & Iverson, 1993). Similar results have been found with speakers of other languages lacking an /r/-/l/ contrast (Korean: Ingram & Park,

1998; Cantonese: Henly & Sheldon, 1986). For example, Korean has also only one liquid, but in Korean it can occur in the coda as well as the onset position. The Korean initial liquid is phonetically realized as a flap /r/ and found only in loan words. In the coda position, the Korean liquid appears as a lateral approximant /l/.

Japanese speakers' ability to perceive /r/ and /l/ depends on position within a word (Sheldon & Strange, 1982; Pisoni et al., 1994). It is reported that the contrast in the initial cluster consonant is the most difficult /r/-/l/ contrast for Japanese listeners to distinguish (Goto, 1971; Gillette, 1980; Sheldon and Strange, 1982; Lively et al., 1993; Ingram et al., 1998). Ingram et al. (1998) investigated /r/-/l/ contrast in several positions by comparing the perceptions of Japanese and Korean listeners. Korean speakers had more trouble with distinguishing the contrast of the word-initial singleton than the cluster or medial position. On the other hand, Japanese speakers found it hardest to discriminate /r/ and /l/ in consonant clusters, and easiest in word-final position (Ingram et al., 1998). In a related study, Henly and Sheldon (1986) investigated the perceptual abilities of native speakers of Cantonese. They demonstrated that identification of /r/ and /l/ in final position and in initial consonant clusters was more difficult than identification of /r/ and /l/ in word-initial position and medial position. These cross-language phonetic studies indicate that the phonology of the listeners' native language is responsible for the differences in performance observed between the two groups of listeners. Similar sounds in the listener's native language influence the perception of L2 sounds.

From these researches, it can be predicted that consonant clusters are the hardest environment for native Japanese speakers to perceive /r/ and /l/, and word-final position is the easiest. But the perceptual relationship between English /r/ and /l/ and the Japanese /r/ is still uncertain. Aoyama et al. (2004) addressed the fact that English /l/ is more strongly assimilated into the Japanese /r/ category than is the English /r/, and this makes the English /r/ easier to learn. Other researchers (Iverson et al., 1995) reported the opposite pattern of learning, but agreed with Aoyama that there are cue biases, which are strongly assimilated into the Japanese /r/ category. The following acoustical research indicates that native and non-native speakers might rely on different acoustic cues when perceiving the same sound contrasts. Iverson et al. (2003) focused on native English adults and native Japanese adults and tested their perceptual abilities of /ra/ and /la/ tokens. His research has shown that the native Japanese rely more on the F2 than the F3 formant frequency in distinguishing the English /r/–/l/ contrast. On the other hand, the native

English adults were sensitive to F3 differences in the perceptual process of categorizing the /r/ and the /l/. Iverson et al. (2003) also investigated the secondary cues (closure durations and transitions) which are important to L2 phoneme learning, but were not predicted as the perceptual interference account. They claimed that the secondary acoustic cues improve L2 learners' identification performance. This research suggests that native and non-native speakers might rely on different acoustic cues when perceiving the same sound contrasts.

The cross-linguistic differences in phonetic realizations are central to the present study. However, there are not enough real data yet to determine what kind of acoustic cue the Japanese adults use when distinguishing /r/-/l/ contrast in the clusters in word-initial position. The phonetic studies of this phenomenon will help us to understand how novel phonemic contrasts are acquired in second language learning. In this study, I attempt to describe an L2 learning task by obtaining distributions of acoustic measures which are used by many phonology and phonetic researchers, and examine the reliability of a novel measure by using /r/-/l/ contrast in the clusters in word-initial position.

3. Perceptual testing techniques

A variety of processes and multiple levels of structure exist in the perception of L2 speech. Therefore, it is insufficient to investigate the difference between perception of native listeners and L2 listeners. This is because testing methods would make it impossible to determine if the score gains observed in a longitudinal research were due to an effect of test familiarization or to the result of perceptual training. The cross-language speech perception abilities of L2 listeners were often measured by using an identification test and/or a discrimination test with non-native words that were thought to be difficult for L2 learners to discriminate. Identification tests are traditional testing method used by phonetic researchers for adult subjects. Discrimination tests are commonly introduced in cross-language speech perception research because they do not require phonological knowledge. It is considered to be appropriate to design L2 perceptual research using one or both of them. Most phonetic researchers employ a discrimination test using the same non-native words as the identification test which was administered in predicting the L2 perceptual learning. However, Flege et al. (1999) developed a new research design called categorial discrimination test (CDT). They claimed that CDT involves characteristics of both an identification and a discrimination test. This study employs these three kinds of

perception tests to investigate the specific type of phonetic perception by Japanese listeners, which is the English /r/-/l/ contrast.

3.1 Identification tests

There are approximately three types of research methods in L2 speech perception, which are behavioral, computational, and neurophysiological methods. Identification tests are categorized as the most direct behavioral method for evaluating the perception of L2 sounds. Behavioral experiments are based on conscious decisions made by a participant. An identification test, a discrimination test, a similarity rating, etc. all belonged to this category. These types of testing techniques help to provide a basic description of how L2 listeners perceive and categorize L2 speech sounds, and scores may show patterns of perceptual confusion.

Categorical perception studies often employ identification scores which present the stimuli of word-initial singleton /r/ and /l/ tokens spoken by native English speakers to show the category boundary for a categorical perception of two consonants tested. For example, Bradlow et al. (1999) used the identification test in their training program at ATR Human Information Processing Research Laboratories in Kyoto to investigate the long-term retention of learning in both perception and production of /r/-/l/ contrast, and indicated that the knowledge gained during perceptual learning of /r/ and /l/ could be transferred to the production domain. Most of studies focused on L2 learners' perception observed the phenomenon that inexperienced Japanese learners of English often misidentify word-initial English liquids in a two-alternative forced-choice test. (Goto, 1971; Mochizuki, 1981; Sheldon and Strange, 1982; Logan et al., 1991; Yamada et al., 1992; Lively et al., 1993; Takagi, 1993; Lively et al., 1994). These inexperienced Japanese subjects are examined to identify a correct rate of 69%.

However, there are three serious methodological problems in identification testing. First, it is concerned with the type of labels that are used to categorize L2 sounds. An identification test should provide response alternatives such as phonetic symbols, spelling or other L2 sound categories. Most of these are not well known by inexperienced L2 learners. Second problem is that many phonetic labels should be offered to get the applicable results. Too many labels may lead listeners to show biased responses, and too few labels may restrict the choices of L2 learners. Most of the perception experiments employ the two-alternative forced-choice (2AFC) method, which provides two response alternatives. In

the 2AFC paradigm, the subject is presented with two alternatives in each trial in which the stimulus is presented. The subject is forced to choose the location or interval at which the stimulus occurred. One disadvantage of 2AFC is that participants can guess the correct answer in 50% of instances.

Third, and the most serious problem, is the influence of stimuli's familiarity. Flege et al. (1996) hypothesized that Japanese speakers' accuracy in identifying /r/ and /l/ might be influenced by the familiarity of the two English words demonstrated as a minimal pair. This phenomenon is supported by the research by Yoshida et al. which revealed that inexperienced Japanese speakers tended to respond with the more familiar of the two tokens of a minimal pair when they knew both tokens of the pair. In the research of Yamada et al. (1992), Japanese subjects showed a positive correlation between lexical familiarity ratings and the percentage of correct identifications of English /r/ and /l/ tokens, and the bias was significant in the scores of subjects who identified English liquids poorly (80% correct). The results observed in certain minimal pairs were: the Japanese subjects identified /r/ far more often in 'red' than in 'rook' (83% vs 29% correct) ; inexperienced Japanese subjects tended to misidentify liquids in a non-word that had a real-word minimal pair than to misidentify liquids found in a real word that was paired with another real word (Yoshida et al., 1988; Yoshida and Seya, 1990). These experiments suggest that less experienced L2 learners are strongly influenced by the lexical background of words used as perceptual stimuli. Therefore, to overcome this problem, most L2 researchers promote both the identification and discrimination tests during one experiment.

3.2 Discrimination tests

Discrimination refers to a listener's ability to distinguish perceptually between two phones. Subjects use phonetic detail to judge the stimuli in discrimination tests, and they rely on stored phonological features to answer in identification tests. There is a significant difference in results of identification and discrimination tests between an inexperienced and untrained Japanese group who had lived in the United States for a short period (less than seven months), and an experienced and trained Japanese group who had been in the United States for 18–48 months (Best and Strange, 1992). The experienced group was more assimilated to a native English group in labeling and discriminating English contrasts than the inexperienced group of Japanese subjects.

There are several types of discrimination tests, each classified by the number of stim-

uli presented in the test.

a. Same/different (AX) tests

In an AX discrimination test, the participants will hear two stimuli of the contrastive minimal pair on each trial. Subjects are then asked to decide whether the phonemes in that pair of stimuli were the same or different. Stimuli were presented in AB and BA orders. The AX paradigm has two serious problems as a test of L2 perception. When two sounds are presented for same/different judgments at a short inter-stimulus interval, both sounds can be held in auditory short-term memory, and judgments can be made on the degree of mismatch of the two auditory representations. The problems just described can be reduced, to some extent, by using the identification test on parallel processing.

b. Categorial tests

A categorial test presents multiple stimuli of each target sound contrast (e.g., Gottfried et al., 1985). Participants are encouraged to respond naturally in a categorial procedure. This type of test focuses on how the participant can perceive an auditory difference between a pair of stimuli and categorize as different sounds. A disadvantage of this test is that a listener sometimes judges the stimuli to be "different" because of the inventory that are not phonologically relevant.

c. Triadic tests (AXB or ABX test)

More than two stimuli are presented on each trial in some discrimination tests. In this type of test, the subjects are offered three response alternatives are presented to the subjects (e.g., Gottfried, 1984; Best et al., 1996). Therefore, Flege (1993) called this type as "triadic test". Participants were told to determine if the X token was "the same" as either the A token or the B token. An important advantage of a triadic test is that it forces the participant to store two stimuli in auditory short-term memory while perceiving a third stimulus. This decreases the bias of auditory-based judgments, and promotes discriminations based on the comparison of auditory representations, especially if a long inter-stimulus interval is used. However, if the interval is short, the results that discrimination tests yield are almost always judged from the auditory information stored in long-term memory. In AXB and ABX tests, the stimuli that are stabled in the 'X' position should be compared to instances of two categories, A and B. This reduces task uncertainty. A disadvantage associ-

ated with the use of the ABX (or AXB) format is that participants can guess the correct answer in 50% of instances. However, an advantage of these type of tests is that they decrease response bias, as guessing will yield the same error rate regardless of whether the participant has a propensity to guess that the variable (X) stimulus occurred in the A or the B position. Flege et al. (1999) tried to overcome the disadvantage of ABX (or AXB) tests by including an equal number of change and catch trials. Change trials contained one sound which differed from the remaining two, and catch trials contained three instances of a single category. They reported that by presenting both change trials and catch trials, the phonetic difference between the sound contrasts could be measured without bias.

d. Oddity test

In the experiment that employs this type of test, three stimuli are provided as in triadic tests. The participants are asked to choose a number from '1', '2' or '3' on the answer sheet to indicate the phonologically odd token out, if they heard one, or to circle '0' if they did not hear an odd token out. The same advantages and disadvantages of triadic tests will be found in the oddity test.

3.3 Categorical discrimination test

A categorial discrimination test (CDT), used by Flege et al. (1999), was designed to assess the discrimination rate of the English sound pairs. The critical difference of CDT from other types of perception tests is that the stimuli used in this test are presented by more than three people. How the stimuli are exposed to subjects depends on which type of testing paradigm the researchers use. An oddity discrimination test was employed in the variant of the CDT used for Flege's study. He reported that the CDT was found to yield stable scores across identification and discrimination tests.

4. Experiment

4.1 Identification test (IT)

Subjects

The subjects were 13 native speakers of Japanese university students (8 male, 5 female) living in Nara and neighboring areas. Their mean age was 19 (range: 19-20). The subjects reported having no hearing disability.

Stimuli

The identification test contained 6 trials of 8 tokens ("pray", "play", "frame", "flame", "free", "flee", "fruit", and "flute"), yielding a total of 48 items. Words which had cluster consonants of the contrast /r/ and /l/ at the initial position were used in the test. Tokens read by a male AE speaker (MS1) were played on a portable CD player (VICTOR CDIOSS QW300). Features of speakers' speech sounds are on Tables 1 and 2.

Procedure

The experiment was conducted during 10 minutes of a 90-minute class period. The CD was played on a CD tape player in the quiet classroom. Subjects marked on their answer sheet whether the stimulus they heard was /r/ or /l/. The answer sheet contained columns of the phonemes /r/ and /l/ so that subjects could circle the answer after hearing the target word. The speech token of each trial was chosen at random. The interval between trials was 2.8 seconds.

4.2 Discrimination test (DT)

Subjects

The same 13 subjects (8 male, 5 female) that participated in the identification test

Table 1. Means of formant frequencies of words spoken by three speakers. (MS=male speaker, FS=female speaker)

Word	Formant	MS1	FS1	FS2	Word	Formant	MS1	FS1	FS2
pray	F1(M)	979.82	747.88	729.46	free	F1(M)	920.77	852.7	886.23
	F2(M)	2057.16	2128.99	2559.3		F2(M)	2086.7	2435.93	2494.38
	F3(M)	2899.18	3009.59	3303.74		F3(M)	2924.29	3168	3288.02
play	F1(M)	860.72	876.5	800.21	flee	F1(M)	838.58	760.71	711.78
	F2(M)	1982.17	2228.77	2586.26		F2(M)	2026.94	2485.84	2571.96
	F3(M)	2855.54	3469.93	3260.16		F3(M)	2910.95	3425.23	3341.47
frame	F1(M)	998.61	725.36	798.54	fruit	F1(M)	1077.11	1353.09	1261.74
	F2(M)	1919.7	1978.8	2209.32		F2(M)	2002.99	2444.8	2024.1
	F3(M)	2916.63	3072.5	3131.94		F3(M)	3047.17	3356.26	3135.6
flame	F1(M)	893.53	769.52	752.72	flute	F1(M)	1157.3	1157.81	1149.95
	F2(M)	1847.27	2105.35	2227.26		F2(M)	2059.11	2153.34	2190.94
	F3(M)	2900.86	3131.54	3199.44		F3(M)	3111.03	3340.97	3263.98

Table 2. Durations of words spoken by three speakers.
(MS=male speaker, FS=female speaker)

Word	pray		play		frame		flame	
Phonemes	/p/	/ray/	/p/	/lay/	/f/	/rame/	/f/	/lame/
MS1	85	340	50	312	62	349	65	353
FS1	119	345	145	319	128	555	197	462
FS2	95	499	106	438	115	534	159	454
Word	free		flee		fruit		flute	
Phonemes	/f/	/ree/	/f/	/lee/	/f/	/ruit/	/f/	/ruit/
MS1	82	234	54	237	42	128	33	377
FS1	276	363	225	460	199	442	289	443
FS2	139	436	178	463	118	477	145	423

answered to this test. Their mean age was 19 (range: 19-20). The subjects reported having no hearing disability.

Stimuli

AXB discrimination test was employed in this task. It contained 6 trials (AAB, ABB, ABA, BAA, BAB, BBA) of 4 minimal pairs ("pray-play", "frame-flame", "free-flee", "fruit-flute"), yielding a total of 24 items. Words which had cluster consonants of the contrast /r/ and /l/ at the initial position were used in this test. Tokens were announced by a male AE speaker (MS1) in the test. The CD was played on a portable CD player (VICTOR CDIOSS QW300).

Procedure

The experiment was conducted during 15 minutes of a 90-minute class period. The CD was played on a portable CD player in the classroom. Three sounds were announced in each question. Three numbers (1, 2, 3) were printed on the answer sheet, and listeners had to mark the numbers that were correct. The three items of each trial were spoken by one speaker and were chosen at random. The interval between the three words in each trial was 1.3 seconds and the interval between trials was 2.8 seconds.

4.3 Categorial discrimination test

Subjects

The same 13 subjects (8 male, 5 female) who participated in the identification and the discrimination test took apart in this test. Their mean age was 19 (range: 19-20). The subjects reported having no hearing disabilities.

Stimuli

AXB type was employed in this categorial discrimination test. It contained 6 trials (AAB, ABB, ACC, AAC, BBC, BCC) of 4 minimal pairs ("pray-play", "frame-flame", "free-flee", "fruit-flute"), yielding a total of 24 items. Words which had cluster consonants of the contrast /r/ and /l/ at the initial position were used in this test. Tokens were recorded by one male (MS1) and two female (FS1 and FS2) AE speakers in the test (Table 1). The CD was played on a portable CD player (VICTOR CDIOSS QW300). The three items of each trial were spoken by three different speakers and were chosen at random. The interval between the three words in each trial was 1.3 seconds and the interval between trials was 2.8 seconds.

Procedure

The experiment was conducted during 15 minutes of a 90-minute class period. The CD was played on a portable CD player in the classroom. Three sounds were announced in each question. Three numbers (1, 2, 3) were printed on the answer sheet, and listeners had to mark the numbers which were adequate. The three items of each trial were spoken by one speaker and were chosen at random. The interval between the three words in each trial was 1.3 seconds and the interval between trials was 2.8 seconds.

5. Results

Stability of test results

The effect of test types (CDT, IT or DT) was non-significant [$F(12, 26) = 0.14$; $p > 0.40$]. This result indicated that three tests do not disagree with each other. The interaction between three tests is presented in Fig. 1. A two-way (test type x pair) ANOVA also yielded non-significant effect for pair [$F(3, 6) = 0.79$, $p>0.50$]. However, a two-way (test type x pair) ANOVA yielded a significant main effect for test type [$F(2, 6) = 14.28$, $p<.006$]. This suggested that the two-way interactions between CDT, DT, and IT were different.

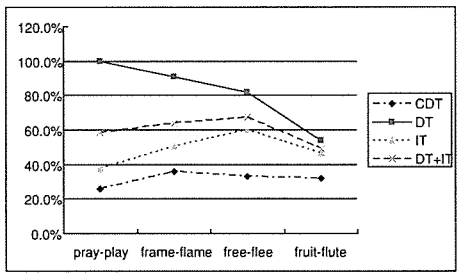

FIG. 1. Means of scores of CDT, DT(discrimination test), IT(identification test) and DT+IT.

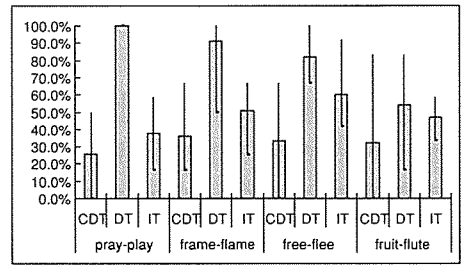

FIG. 2. Means, maximum values and minimal values of scores of CDT, DT, and IT.

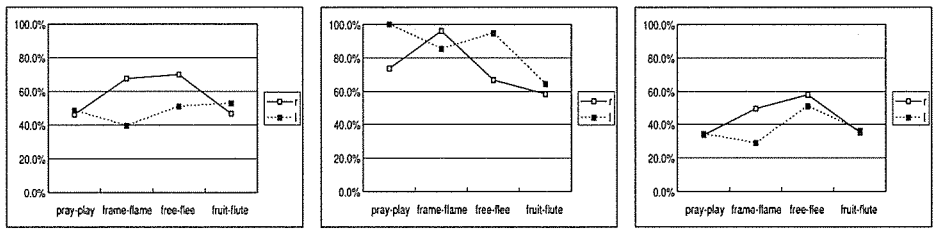

FIG. 3. Interaction plots: IT, DT and CDT data. Comparison between the interaction of item type (/r/ or /l/) with minimal pair for identification and discrimination responses. The graphs illustrate different response profiles for IT, DT and CDT.

The means of the scores of three tests were submitted to Pearson correlation analyses. Significant correlations were obtained for pairs CDT-IT [$r=0.758$, $p<0.05$]. CDT was inversely correlated with DT [$r= -0.30$, $p<0.05$] and a similar correlation was obtained for DT and IT [$r= -0.222$, $p < 0.05$]. The score of CDT test was tightly connected to the result of IT, and therefore, it can be assumed that scores on CDT depended importantly on whether a non-native participant discriminated two distinct sound features when perceiving the stimuli in the trials.

Familiarization with CDT did not seem to have increased test scores. Previous researchers, as hypothesized earlier, indicated that word familiarity effects should operate more strongly in the identification than the discrimination test. In view of the /r/ perception biases observed in the identification task, a comparison of the relative subjective familiarity (Flege et al., 1995) of the /r/-/l/ contrast of the 4 minimal pairs were elicited from CDT. There was a high degree of consensus among the four pairs. In the scores of IT, the /r/ member was rated as more familiar than the /l/ member in two minimal pairs ('frame'-'flame', 'free'-'flee') and opposite in other two minimal pairs ('pray'-'play', 'frute'-

'flute'). Hence any /r/ bias or lexical familiarity effects would be highly correlated and probably indistinguishable in IT, and not in DT and CDT. Taken together, these test scores indicate that the CDT can provide stable test scores to researchers.

Analysis of test scores

In many cases, the difficulty of /r/-/l/ contrast might be predicted based on an assumption of the acoustic-phonetic properties of two sounds and the closest L1 sound, and also empirical evidence concerning how the English liquids are related perceptually to the flap /r/ in Japanese (e.g., Mochizuki, 1981). Results of the identification experiment suggested a dominance hierarchy among the factors contributing to the perceptibility of the /r/-/l/ contrast by Japanese subjects. The pattern of performance for Japanese listeners, who don't possess the L1 category of the L2 contrast, was consistent with considerations of acoustic discriminability, with performance worst in cluster position.

Parallel two-way analyses of variance of IT, DT and CDT tests were conducted, with the four minimal pairs and the token type (/r/ or /l/) as experimental factors. Figure 3, which represents an interaction plot of the two experimental factors, shows the profile of /r/ and /l/ scores across the minimal pairs for IT, DT, and CDT. A significant main effect of minimal pair was found for all three tests [$F(2, 6) = 14.28$, $p<.006$], but the scoring profiles of /r/ and /l/ over the four minimal pairs were quite different for the three tests (see Fig. 3). IT yielded a significant main effect for token type, [$F(7, 3) = 6.71$, $p<0.001$] revealing an /r/ bias for two of the minimal pair contrasts. The opposite effect was found in the case of DT [$F(7, 3) = 8.64$, $p=0.001$] , an /l/ bias for three of the minimal pairs. CDT didn't show a significant effect from token types[$F(7, 3) = 0.24$; $p > 0.20$]. IT also yielded a significant minimal pair by token type interaction, [$F(7, 3) = 1.77$, $p< 0.009$] but no significant interaction effect was found for DT [$F(7, 3) = 0.12$, $p>0.3$].

A significant difference was noted for one contrast ('pray'-'play') which had obtained a high score on DT. On the contrary, CDT and IT scores showed that the 'pray'-'play' was the most difficult pair to perceive in four pairs. The minimal pair 'frame'-'flame' showed a strong /r/ bias on all three tests, but a slight difference between them. These gaps in performance profile across the minimal pairs indicate that IT, DT, and CDT were tapping different perceptual processes or making different task demands of listeners. While a similar pattern of main effects was found in DT to those observed in the IT, the lack of interaction

effects suggested that the two tasks may have invoked different processing strategies on the part of listeners. CDT yielded the both processes of IT and DT. Phonological transfer effects (language-specific learning effects) were diminished in the discrimination experiment. The main effect of language background was preserved, but no interaction between listeners' language of the /r/-/l/ contrast was found. These results were consistent with the hypothesis that underlying perceptual mechanisms differed in IT and DT. It was apparent from IT results that many of the significant higher-order interactions may have been caused by response preferences towards /r/, which varied with the language background of the listener and the position of the target sound. However, such effects were not apparent in the case of the discrimination data.

These results show that CDT can be replaced to the usage of IT and DT in one phonological experiment. The disadvantages of both IT and DT are removed from CDT. However, differences between pairs which are not focused in this research may arise as the result of perceptual learning. Researches which focus on /r/-/l/ contrast in different position or in different cluster of minimal pairs are needed, hence the possibility that native and non-native speakers might rely on different acoustic cues when perceiving the same sound contrasts would give a crucial evidence for further understanding.

6. Discussion

Although the /r/-/l/ contrast had been investigated by many researchers, the new aspects of L2 perception can be found by using novel techniques. In this study, the perception of the /r/-/l/ contrast in initial cluster position was focused on, and the result showed that there was no statistically significant difference between the three types of perceptual tests, which were identification, discrimination, and categorical discrimination test. This result indicates that the CDT yielded stable scores across two test paradigms, and supports the claim (Flege, 1995) that L2 learners are capable of establishing new phonetic categories for L2 sound contrast. This finding is consistent with the view that the CDT would be appropriate for use in a perception study of L2 speech learning. L2 researchers are principally interested in whether or how the listeners were able to perceive the L2 phonemes which are non-existent in L1 speech. However, in order to get a stable result from the L2 perception experiment, more researches of L2 sound contrast may be necessary.

References

Aoyama, K., Flege, J. E., Guion, S. G., Yamada, R. A., and Yamada, T. (2004) Perceived phonetics dissimilarity and L2 speech learning: the case of Japanese /r/ and English /l/ and /r/, *Journal of Phonetics*, 32, 233-250.

Best, C. T. and Strange, W. (1992) Effects of phonological and phonetic factors on cross-language perception of approximants, *Journal of Phonetics*, 20, 305–330.

Best, C. T., Faber, A., and Levitt, A. (1996) Assimilation of non-native vowel contrasts to the American English vowel system, *Journal of the Acoustical Society of America*, 99, 2602.

Bradlow, A. R., Yamada, R. A., Pisoni, D. B., and Tohkura, Y. (1999) Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production, *Perception & Psychophysics*, 61, 5, 977-985.

Broselow, E. and Finer, D. (1991) Parameter setting in second language phonology and syntax, *Second Language Research*, 7, 35-59.

Eckman, F. R. and Iverson, G. K. (1993) Sonority and markedness among onset clusters in the interlanguage of ESL learners, *Second Language Research*, 97, 234-252.

Flege, J. E. (1993) Production and perception of a novel, second-language phonetic contrast, *Journal of the Acoustical Society of America*, 93, 1589-1608.

Flege, J. E. (1995) Second language speech learning theory, findings, and problems, *Speech Perception and Linguistic Experience: Issues in Cross-Language Research 3-52*, Timonium, MD, 233-276.

Flege, J. E., Takagi, N. and Mann, V. (1996) Lexical familarity and English language experience affect Japanese adults' perception of /r/ and /l/, *Journal of the Acoustical Society of America*, 99, 1161-1173.

Flege, J. E., MacKay, I. R. A. and Meador, D. (1999) Native Italian speakers' perception and production of English vowels, *Journal of the Acoustical Society of America*, 106, 2973-2987.

Gillette, S. (1980) Contextual variation in the perception of L and R by Japanese and Korean speakers, *Minn. Papers Ling. Philos. Lang.* 6, 59-72.

Goto, H. (1971) Auditory perception by normal Japanese adults of the sounds 'l' and 'r', *Neuropsychologia*, 9, 317-323.

Gottfried, T. L. (1984) Effects of consonant context on the perception of French vowels. *Journal of Phonetics*, 12, 91–114.

Gottfried. T. R., Jenkins J. J. and Strange. W. (1985) Categorial discrimination of vowels produced in a syllable context and in isolation, *Bulletin of the Psychonomic Society*, 23.10 I-I 04.

Henly, E. and Sheldon, A. (1986) Duration and context elects on the perception of English /r/ and /l/: a comparison of Cantonese and Japanese speakers, *Language Learning*, 36, 505-521.

Ingram, J. C. and Park, S.-G. (1998) Language, context, and speaker effects in the identification and discrimination of English /r/ and /l/ by Japanese and Korean listeners, *Journal of the Acoustical Society of America*, 103, 1161–1174.

Iverson, P. and Kuhl, P. K. (1995) Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling, *Journal of the Acoustical Society of America*, 97, 553-562.

Iverson, P. K., Yamada, R. A., Diesch, E., Tohkura, Y., Kettermann, A. and Siebert, C. (2003) A perceptual interference account of acquisition difficulties for non-native phonemes, *Cognition*, 87, B47-B57.

Lively, S. E., Logan, J. S., and Pisoni, D. B. (1993) Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories, *Journal of the Acoustical Society of America*, 94, 1242-1255.

Lively, S., Yamada, R., Tohkura, Y, and Yamada, T. (1994) Training Japanese listeners to identify English /r/ and /l/, III. Long-term retention of new phonetic categories, *Journal of the Acoustical Society of America*, 96, 2076–2087.

Logan, J. S., Lively, S. E., and Pisoni, D. B. (1991) Training Japanese listeners to identify English /r/ and /l/: A first report, *Journal of the Acoustical Society of America*, 89, 874-886.

Mackain, K. S., Best, C. T., and Strange, W. (1981) Categorical perception of English /r/ and /l/ by Japanese bilinguals, *Applied Psycholinguistics*, 2, 369-390.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J. and Fujimura, O. (1975) An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English, *Perception & Psychophysics*, 18. 5, 331-340.

Mochizuki, M. (1981) The identification of /r/ and /l/ in natural and synthesized speech, *Journal of Phonetics*, 9, 283-303.

Pisoni, D. B., Lively, S. E., and Logan, J. S. (1994) Perceptual learning of nonnative speech

contrasts: Implications for theories of speech perception. In J. C. Goodman & H. C. Nusbaum (eds.), *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words* (pp. 121-166), Cambridge, MA: MIT Press.

Sato, C. (1984) Phonological processes in second language acquisition: another look at interlanguage syllable structure, *Language Learning*, 34, 43-57.

Sheldon, A. and Strange, W. (1982) The acquisition of /r/ and /l/ by Japanese learners of English: evidence that speech production can precede perception, *Applied Psycholinguistics*, 3, 243-261.

Takagi, N. (1993) Perception of American English /r/ and /l/ by adult Japanese learners of English: A Unified View, *Unpublished Ph.D. dissertation, Univ. of California-Irvine.*

Takagi, N. and Mann, V. A. (1995) The limits of extended naturalistic exposure on the perceptual mastery of English /r/ and /l/ by adult Japanese learners of English, *Applied Psycholinguistics*, 16. 379–405.

Yamada, R. A., and Tohkura, Y. (1992) The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners, *Perception & Psychophysics*, 52, 376-392.

Yoshida, K., Saito, Y., Matsumoto, Y., Hayakawa, Y., Nemoto, C., Tanaka, M., and Iwahashi, Y. (1988) Phoneme perception and lexical knowledge in second language acquisition, *Sophia Linguistica*, 23/24, 237–242.

Yoshida, K., and Seya, H. (1990) The use of the lexical strategy in the perception of English consonants by Japanese returnee children, *Maintenance of Foreign Languages by Returnee Children* (Japan Overseas Education Services, Tokyo), Vol. 2, pp. 29–41.